

LAPORAN UJIAN TENGAH SEMESTER
STATISTIKA DESKRIPTIF

“MEMVISUALISASIKAN DATA GAJI TAHUNAN PEGAWAI IT DI EROPA TAHUN
2020”



Disusun oleh:

Mukhamad Ikhsanudin (082011633086)

Mata Kuliah:

Statistika Deskriptif

Dosen Pengampu:

Eto Wuryanto, Drs., DEA

PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA
2021

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tujuan dari survei dalam data ini adalah untuk mempelajari nilai kompetitif dari keahlian untuk spesialis IT yang bergantung pada pengalaman tahunan, posisi, bahasa, dan lainnya. Survei gaji anonim pada data ini telah dilakukan setiap tahun sejak 2015 di antara spesialis IT Eropa dengan fokus yang lebih kuat di Jerman. Tahun 2020 ini terdapat 1238 responden secara sukarela berpartisipasi dalam survei. Data yang telah dipublikasikan oleh akun bernama Parul Pandey di website kaggle.com ini bisa digunakan secara umum. Dataset ini berisi informasi tentang pola gaji di antara para profesional IT di wilayah Eropa dan juga pilihan yang baru.

Dalam analisis ini, "Gaji" berarti gaji tahunan bruto dalam ribuan euro (disingkat menjadi k), termasuk semua pajak. Isi dari data ini diantaranya :

- a. Usia (Age)
- b. Jenis kelamin (Gender)
- c. Kota (City)
- d. Bahasa (Main language at work)
- e. Cuti tahunan (Number of vacation days)
- f. Gaji (Yearly salary)
- g. Posisi (Position)
- h. Status pegawai (Freelance / Fulltime)
- i. Pengalaman kerja (Years of experience)
- j. Senioritas (Seniority level)
- k. Dampak Corona terhadap pegawai (Losing job after Coronavirus)

Dari dataset yang sangat lengkap tersebut, kita bisa membuat visualisasi data yang sangat beragam. Maka dari itu, untuk mempersingkat pembuatan dan mempermudah pemahaman, penulis hanya menggunakan beberapa data. Selain itu, data yang tidak digunakan juga sangat bervariasi sehingga dikhawatirkan akan membuat visualisasi data kurang jelas.

1.2 Tujuan Pengolahan Data

- 1.2.1 Mengetahui perbandingan rata-rata gaji tahunan pegawai IT di Eropa pada tahun 2020 berdasarkan beberapa faktor lain
- 1.2.2 Mengetahui hubungan antara gaji pegawai dengan data lainnya
- 1.2.3 Mengetahui penampakan data secara visual dengan berbagai bentuk
- 1.2.4 Mengetahui perbandingan antardata selain gaji tahunan yang masih berkaitan dengan pegawai

BAB II

KAJIAN PUSTAKA

2.1 Visualisasi Data (*Data Visualization*)

Visualisasi data mengikutkan pembuatan dan kajian dari representasi visual dari data, artinya "informasi yang telah diabstraksikan dalam bentuk skematis, termasuk atribut atau variabel dari unit informasi.

Tujuan utama dari visualisasi data adalah untuk mengkomunikasikan informasi secara jelas dan efisien kepada pengguna lewat grafik informasi yang dipilih, seperti tabel dan grafik. Visualisasi yang efektif membantu pengguna dalam menganalisis dan penalaran tentang data dan bukti.

2.2 R

R (juga dikenal sebagai GNU S) adalah bahasa pemrograman dan perangkat lunak untuk analisis statistika dan grafik. R dibuat oleh Ross Ihaka dan Robert Gentleman di Universitas Auckland, Selandia Baru, dan kini dikembangkan oleh R Development Core Team, di mana Chambers merupakan anggotanya. R dinamakan sebagian setelah nama dua pembuatnya (Robert Gentleman dan Ross Ihaka), dan sebagian sebagian dari permainan nama dari S.

Bahasa R kini menjadi standar de facto di antara statistikawan untuk pengembangan perangkat lunak statistika, serta digunakan secara luas untuk pengembangan perangkat lunak statistika dan analisis data.

R menyediakan berbagai teknik statistika (permodelan linier dan nonlinier, uji statistik klasik, analisis deret waktu, klasifikasi, klusterisasi, dan sebagainya) serta grafik. R, sebagaimana S, dirancang sebagai bahasa komputer sebenarnya, dan mengizinkan penggunaannya untuk menambah fungsi tambahan dengan mendefinisikan fungsi baru. Kekuatan besar dari R yang lain adalah fasilitas grafiknya, yang menghasilkan grafik dengan kualitas publikasi yang dapat memuat simbol matematika. R memiliki format dokumentasi seperti LaTeX, yang digunakan untuk menyediakan dokumentasi yang lengkap, baik secara daring (dalam berbagai format) maupun secara cetakan.

Meskipun R memiliki antarmuka baris perintah (command line interface), terdapat beberapa antarmuka pengguna grafis pihak ketiga, seperti RStudio, lingkungan pengembangan yang terintegrasi, dan Jupyter, antarmuka notebook.

2.3 R-Studio

RStudio adalah lingkungan pengembangan terintegrasi (IDE) sumber terbuka dan gratis untuk R, bahasa pemrograman untuk komputasi statistik dan grafik. RStudio didirikan oleh JJ Allaire, pencipta bahasa pemrograman ColdFusion . Hadley Wickham adalah Kepala Ilmuwan di RStudio.

RStudio tersedia dalam dua edisi: RStudio Desktop, di mana program dijalankan secara lokal sebagai aplikasi desktop biasa; dan RStudio Server, yang memungkinkan mengakses RStudio menggunakan browser web saat sedang berjalan di server Linux jarak jauh. Distribusi prepackaged Desktop RStudio tersedia untuk Windows, macOS, dan Linux .

RStudio tersedia dalam edisi open source dan komersial dan berjalan di desktop (Windows, macOS, dan Linux) atau di browser yang terhubung ke RStudio Server atau RStudio Server Pro (Debian, Ubuntu, Red Hat Linux, CentOS, openSUSE dan SLES).

RStudio sebagian ditulis dalam bahasa pemrograman C ++ dan menggunakan kerangka kerja Qt untuk antarmuka pengguna grafisnya . Persentase kode yang lebih besar ditulis dalam Java. JavaScript juga di antara bahasa yang digunakan.

Pekerjaan di RStudio dimulai sekitar Desember 2010, dan versi beta publik pertama (v0.92) secara resmi diumumkan pada Februari 2011. Versi 1.0 dirilis pada 1 November 2016. Versi 1.1 dirilis pada 9 Oktober 2017.

Pada bulan April 2018 diumumkan RStudio akan menyediakan dukungan operasional dan infrastruktur untuk Ursa Labs. Ursa Labs akan fokus membangun runtime sains data baru yang didukung oleh Apache Arrow.

2.4 Package

Package adalah sekumpulan fasilitas fungsi yang dapat digunakan oleh *user* untuk melakukan tugas. *Package* yang digunakan dalam proses ini yaitu :

ggplot2 - These functions provides tools to help you make visualization data with ggplot2
plotly - These functions provides tools to help you to read the visualized data

BAB III

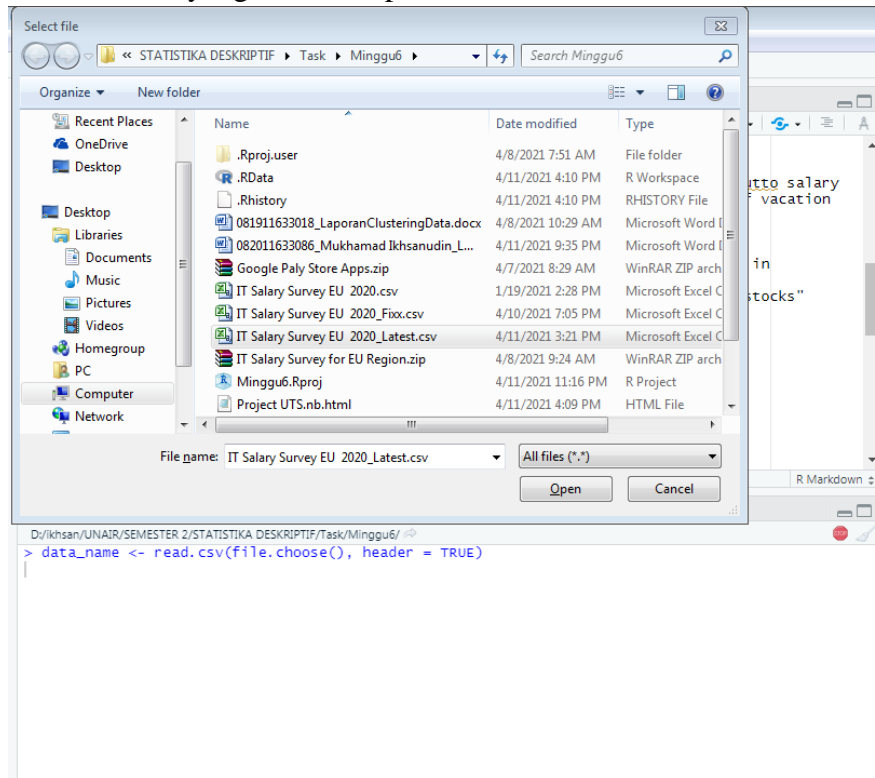
PEMBAHASAN

3.1 Prosedur Mengimport dataset IT Salary Survey 2020

3.1.1 Memasukkan data ke dalam R-Studio

```
data_name <- read.csv(file.choose(), header = TRUE)
```

3.1.2 Memilih data yang akan diimport



3.1.3 Melihat data yang sudah diimport

```
> View(data_name)
```

3.1.4 Mengecek data yang akan digunakan

```
> str(data_name)
spec_tbl_df [1,105 x 9] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Age           : num [1:1105] 26 26 29 28 37 32 37 24 29 35 ...
 $ Gender        : chr [1:1105] "Male" "Male" "Male" "Male" ...
 $ Experience_Years : num [1:1105] 5 7 12 4 17 5 6 5 8 15 ...
 $ Year_Salary    : num [1:1105] 80000 80000 120000 54000 62000 76000
 57000 65000 56000 95000 ...
 $ Bonus_Stocks   : num [1:1105] 5000 0 120000 0 0 5000 0 0 0 0 ...
 $ Vacation_Days   : num [1:1105] 30 28 30 24 29 30 24 27 28 30 ...
 $ Work_Language   : chr [1:1105] "English" "English" "English" "English" ...
 $ Company_Type    : chr [1:1105] "Product" "Product" "Product" "Startup" ...
 $ Fired_By_Corona : chr [1:1105] "No" "No" "Yes" "No" ...
 - attr(*, "spec")=List of 3
```

```

..$ cols :List of 9
.. ..$ Age : list()
.. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Gender : list()
.. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Experience_Years: list()
.. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Year_Salary : list()
.. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Bonus_Stocks : list()
.. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Vacation_Days : list()
.. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Work_Language : list()
.. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Company_Type : list()
.. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Fired_By_Corona : list()
.. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
..$ default : list()
..- attr(*, "class")= chr [1:2] "collector_guess" "collector"
..$ skip : int 1
..- attr(*, "class")= chr "col_spec"

```

3.2 Melihat Persebaran Data Gaji Tahunan Pegawai IT di Eropa pada 2020 Berdasarkan Pengalaman Kerja dan Kondisi Setelah Adanya *Coronavirus*

3.2.1 Menginstall *packages*

```

install.packages(ggplot2)
install.package(plotly)

```

3.2.2 Menggunakan *library* pada *packages* yang sudah diinstall

```

library(ggplot2)
library(plotly)

```

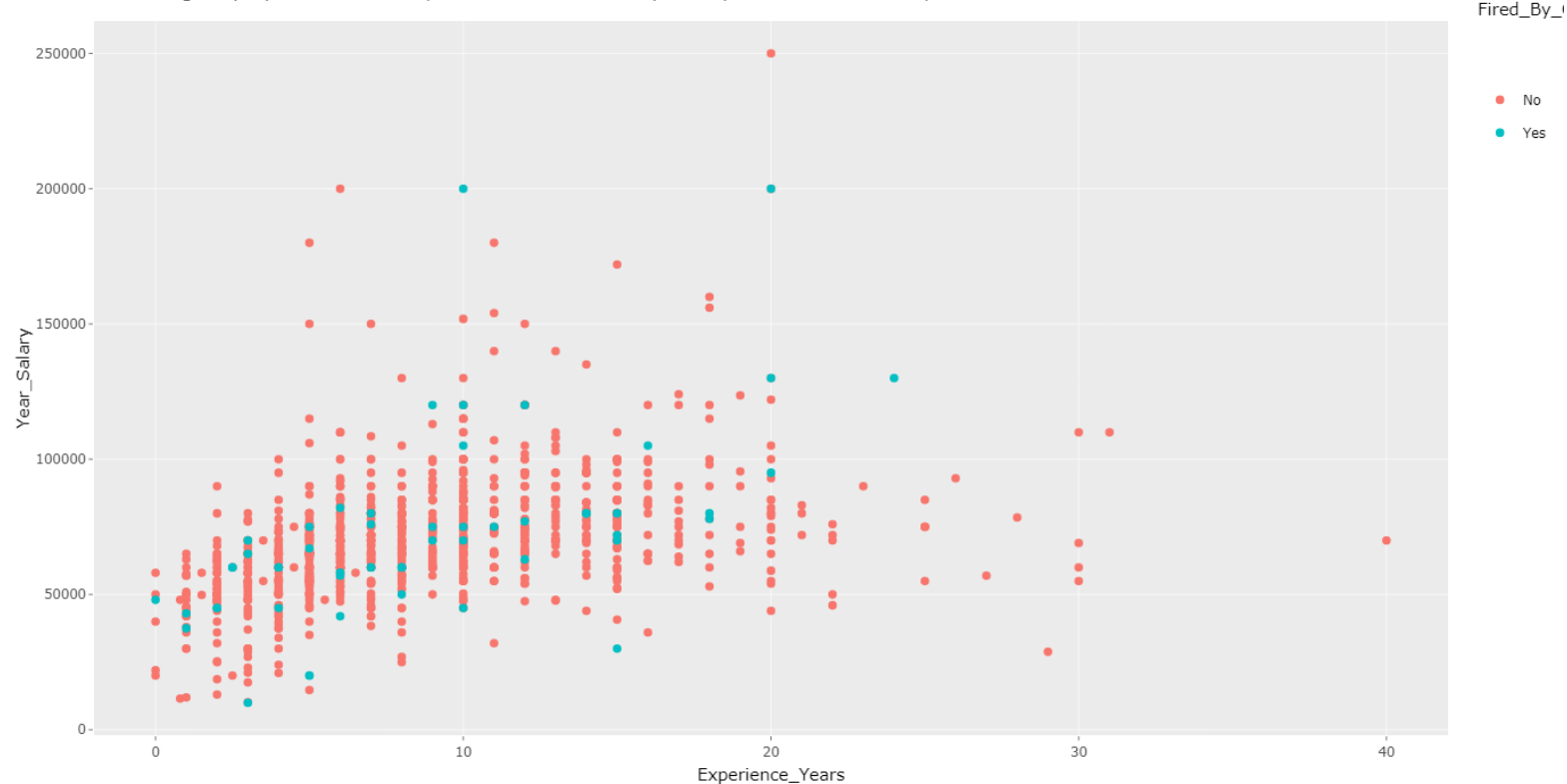
3.2.3 Membuat grafik multivariat (*multivariate graphs*) berupa point (titik) yang merupakan interpretasi dari tiga data (gaji, pengalaman kerja, dan pemecatan)

```

ggplotly(ggplot(IT_Salary, aes(x = Experience_Years,
                               y = Year_Salary,
                               color = Fired_By_Corona)) +
  geom_point() +
  labs(title = "The Firing Employees Relationship based on Their Yearly Salary and
Total Work Experience"))

```

Dari grafik, kita bisa melihat pegawai yang dipecat dan tidak berdasarkan warna, lalu persebaran data terbanyak terdapat pada gaji antara 0 – 150000 Euro dan pengalaman kerjanya 0 – 20 tahun. Namun terdapat outlier di gaji 250000 dan pengalaman kerja 40 tahun



3.3.1 Menginstall packages

3.3.2 Menggunakan *library* pada *packages* yang sudah diinstall

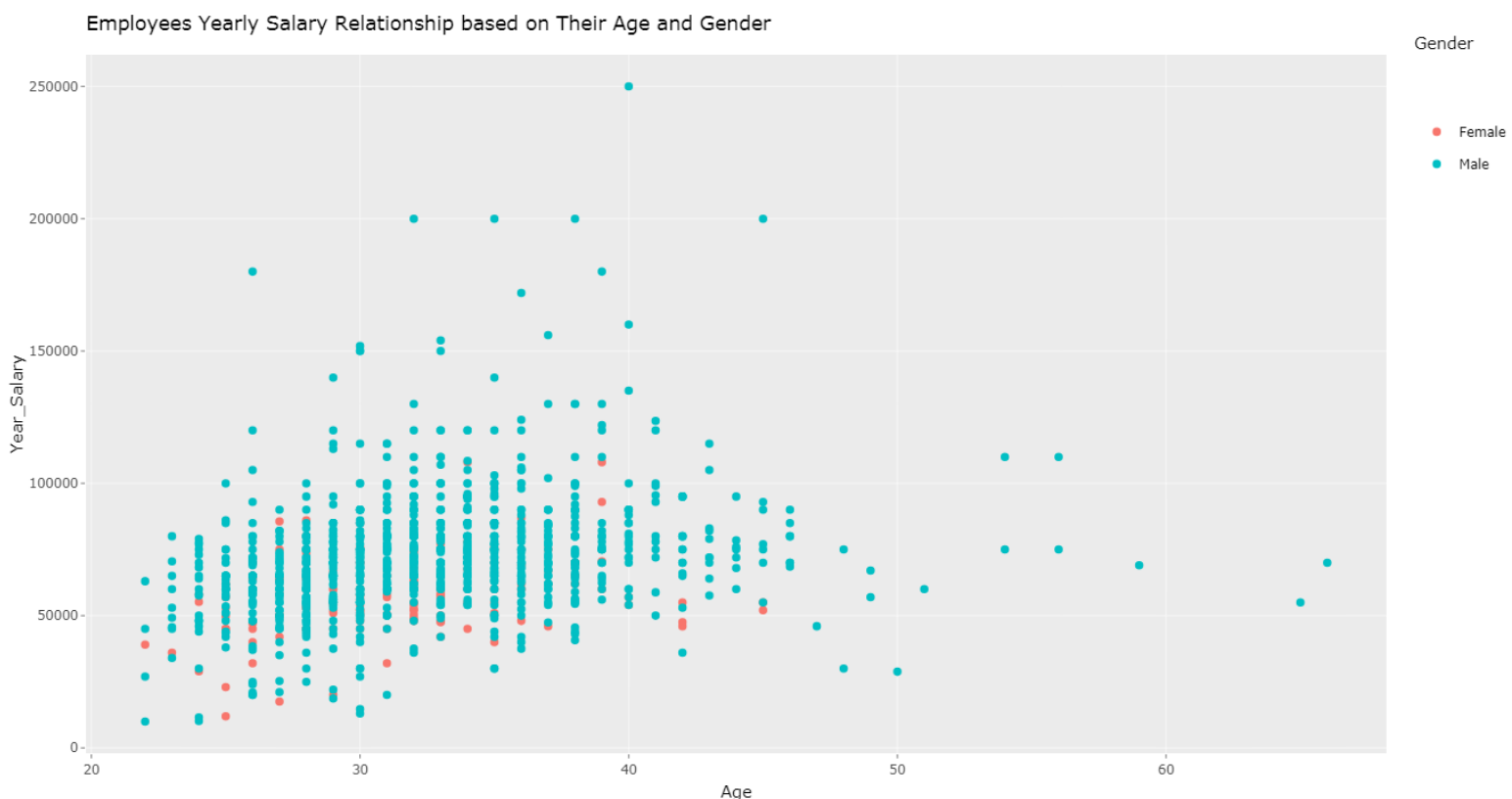
3.3.3 Membuat grafik multivariat (*multivariate graphs*) berupa point (titik) yang merupakan interpretasi dari tiga data (gaji, umur, dan jenis kelamin)

[illegible]

```
labs(title = "Employees Yearly Salary Relationship based on Their Age and Gender"))
```

3.3.4 Menginterpretasi persebaran data gaji tahunan pegawai IT di Eropa berdasarkan umur dan jenis kelamin

Dari grafik, kita bisa melihat perbedaan antara laki-laki dan perempuan berdasarkan warna. Lalu persebaran umur yang terbanyak terdapat pada umur 20 – 40 tahun. Ada outlier di umur >60



3.4 Melihat Persebaran Data Bahasa Kerja Tahunan Pegawai IT di Eropa pada 2020 Berdasarkan Bonus dan Saham serta Jenis Kelamin

3.4.1 Menginstall packages

```
install.packages(ggplot2)  
install.packages(plotly)
```

3.4.2 Menggunakan library pada packages yang sudah di install

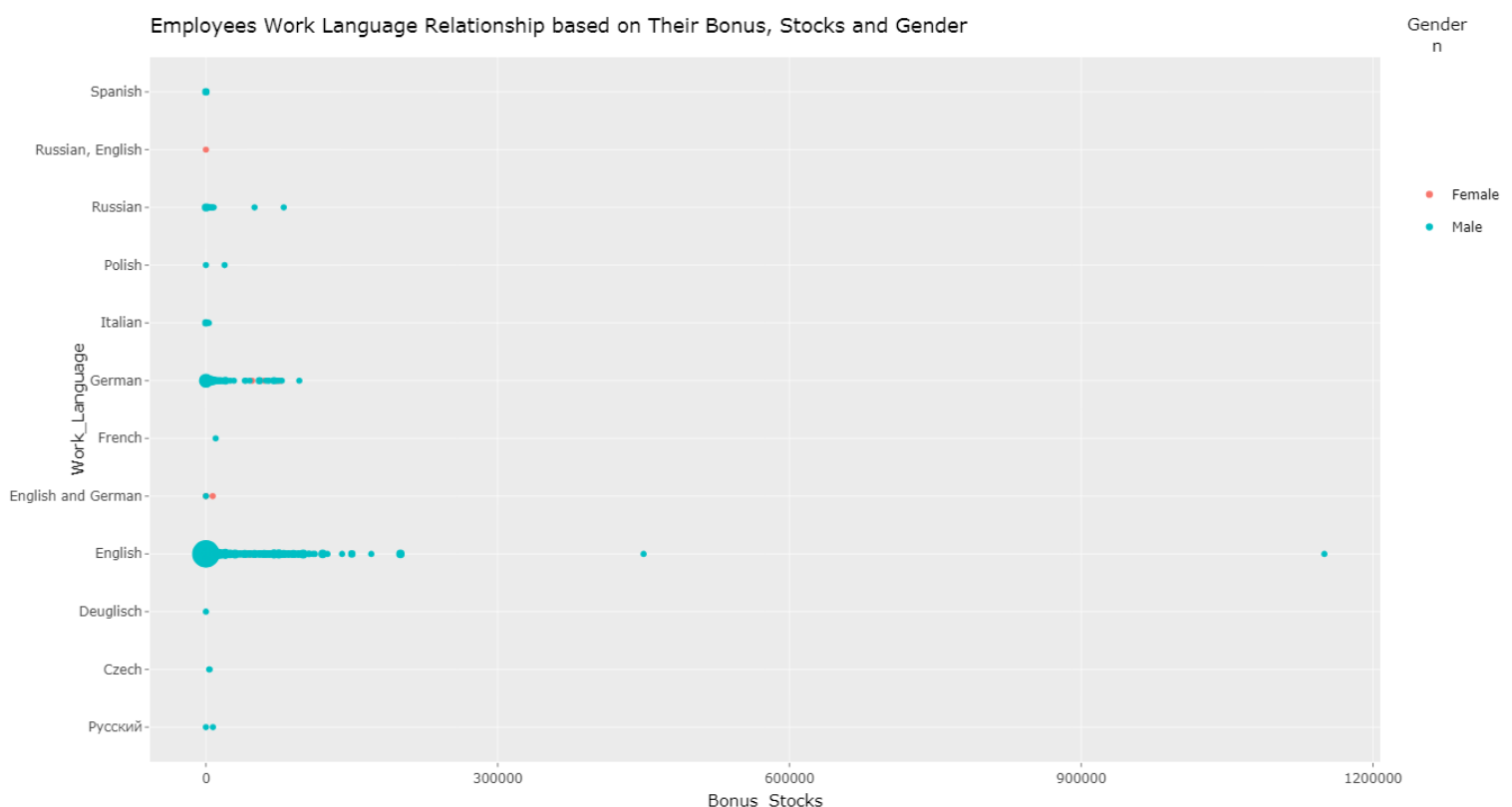
```
library(ggplot2)  
library(plotly)
```

3.4.3 Membuat grafik multivariat (*multivariate graphs*) berupa point(titik) yang merupakan interpretasi dari tiga data (bahasa kerja, bonus dan saham, serta jenis kelamin)


```
ggplotly(ggplot(IT_Salary) +
  geom_count(aes(x = Bonus_Stocks,
    y = Work_Language,
    color = Gender)) +
  labs(title = "Employees Work Language Relationship based on Their Bonus,
    Stocks and Gender"))
```

3.4.4 Menginterpretasi persebaran data bahasa kerja pegawai IT di Eropa pada tahun 2020 berdasarkan bonus dan saham serta jenis kelaminnya

Dari grafik kita bisa mengetahui bahwa bahasa pegawai IT di Eropa yang terbanyak bahasa Inggris, lalu Jerman. Selain itu, terdapat outlier bonus di antara 300000-600000 dan di antara 900000-1200000



3.5 Melihat Histogram Data Gaji pegawai IT di Eropa Tahun 2020 Berdasarkan Jenis Kelamin

3.5.1 Menginstall packages

```
install.packages(ggplot2)
install.packages(plotly)
```

3.5.2 Menggunakan *library* pada *packages* yang sudah di install

```
library(ggplot2)
library(plotly)
```

3.5.3 Membuat grafik bivariate (*bivariate graphs*) berupa histogram yang merupakan interpretasi dari dua data (bahasa kerja, bonus dan saham, serta jenis kelamin)

```
ggplotly(ggplot(IT_Salary, aes(x = Year_Salary)) +  
  geom_histogram(fill = "cornflowerblue",  
    color = "white") +  
  facet_wrap(~Gender, ncol = 1) +  
  labs(title = "Employees Salary Histograms by Gender"))
```

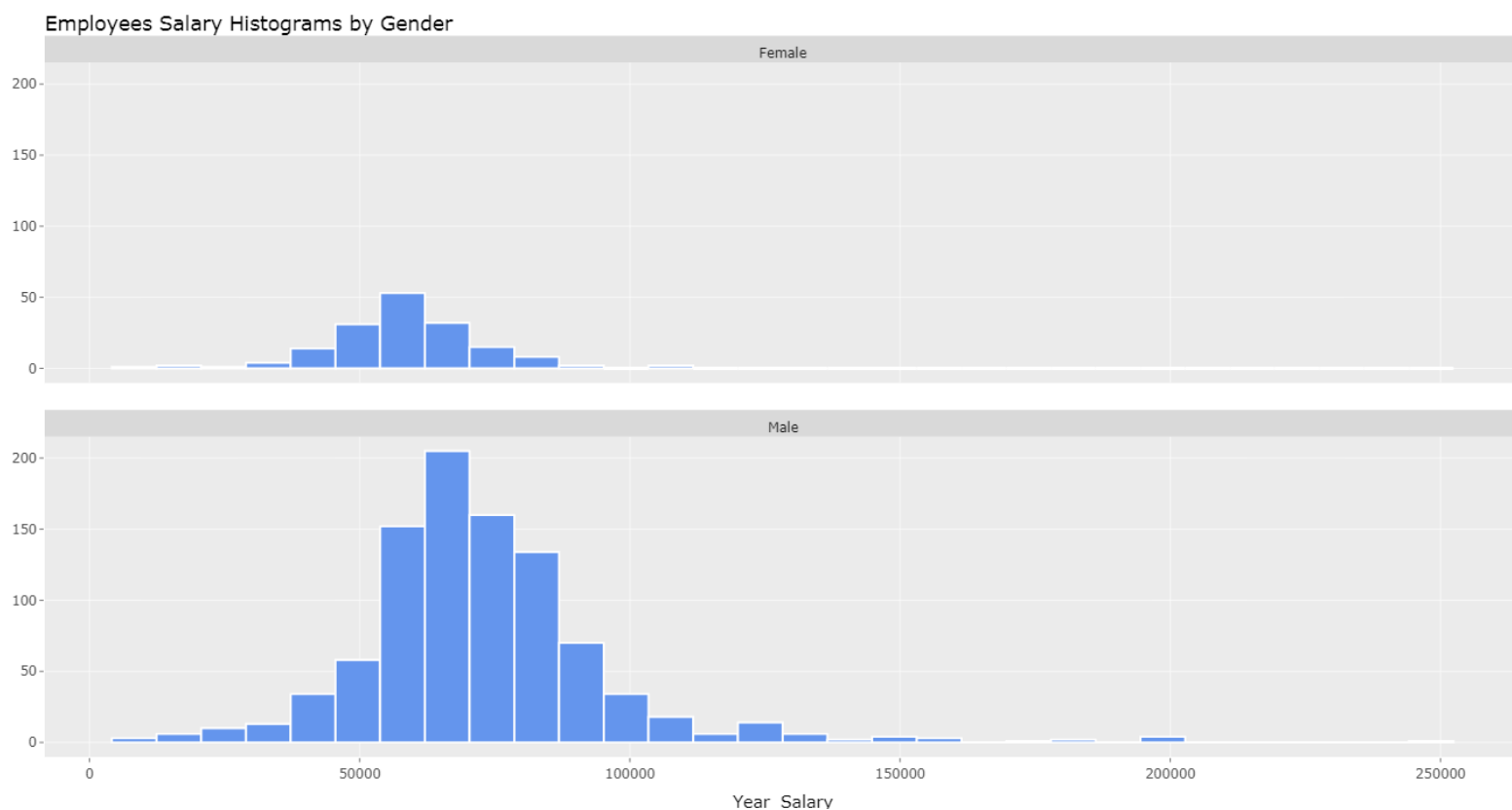
3.5.4 Menginterpretasi histogram data gaji pegawai IT di Eropa pada tahun 2020

berdasarkan jenis kelaminnya

Dari grafik, kita bisa melihat perbedaan antara pegawai laki-laki dan perempuan.

Pada grafik laki-laki, terdapat outlier gaji, sedangkan pada perempuan tidak ada.

Pada grafik perempuan, terdapat 50 pegawai dengan rentang gaji yang sama, sedangkan pegawai laki-laki terdapat ± 200 pegawai.



3.6 Melihat Persebaran Data Pegawai (Gaji, Umur, dan Jenis Kelamin) IT di Eropa pada 2020 Tiap Bahasa Kerja

3.6.1 Menginstall packages

```
install.packages(ggplot2)
```

```
install.packages(plotly)
```

3.6.2 Menggunakan *library* pada *packages* yang sudah di install

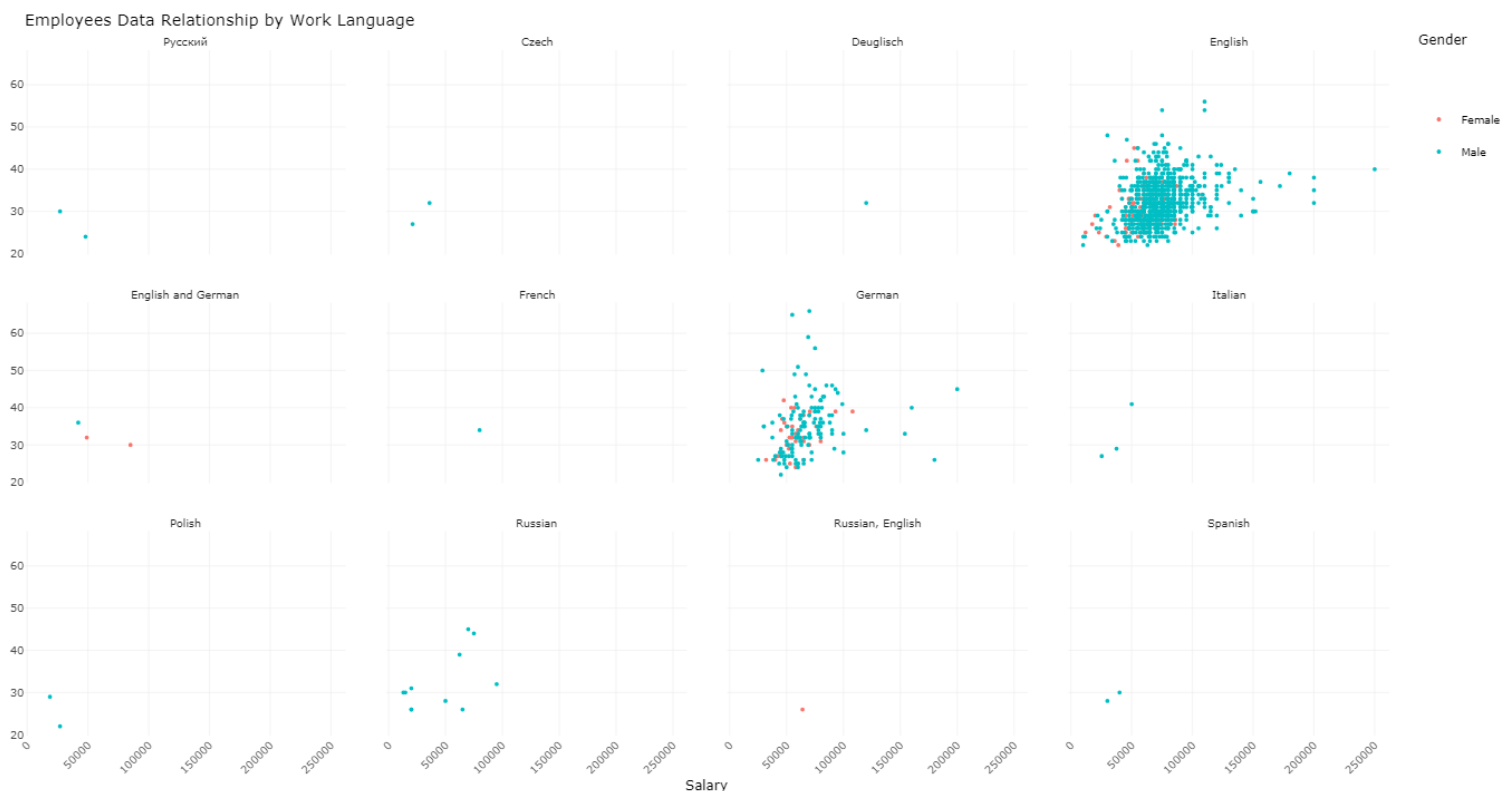
```
library(ggplot2)
library(plotly)
```

3.6.3 Membuat grafik multivariat (*multivariate graphs*) berupa point (titik) yang merupakan interpretasi dari tiga data (gaji, umur, dan jenis kelamin) tiap bahasa kerja

```
ggplotly(ggplot(IT_Salary, aes(x=Year_Salary,
                              y = Age,
                              color = Gender)) +
  geom_point(size = 0.5) +
  facet_wrap(~Work_Language) +
  theme_minimal(base_size = 9) +
  theme(axis.text.x = element_text(angle = 45,
                                     hjust = 1)) +
  labs(title = "Employees Data Relationship by Work Language",
        x = "Salary",
        y = "Age"))
```

3.6.4 Menginterpretasi persebaran data pegawai (gaji, umur, dan jenis kelamin) IT di Eropa pada tahun 2020 berdasarkan bahasa kerja

Pada proses ini, terdapat 12 grafik sesuai banyak bahasa yang didapatkan dari survei. Bahasa terbanyak yaitu bahasa Inggris, lalu bahasa Jerman. Lalu, angka di samping menunjukkan umur pegawai.



BAB IV

PENUTUP

4.1 Kesimpulan

Visualisasi data (*data visualization*) sangat berguna bagi kita. Data yang telah divisualisasikan akan lebih mudah untuk dibaca dan dipahami maksudnya. Namun, dalam memvisualisasikan data, kita juga perlu membuat kerangka tujuannya terlebih dahulu. Dengan kata lain, visualisasi data berhasil jika telah mencapai tujuan tersebut.

Dari pengolahan data yang telah dilakukan, dapat diambil kesimpulan bahwa :

1. Data yang diolah memiliki hubungan antara satu dengan yang lain
2. Visualisasi data memiliki berbagai interpretasi bentuk diagram, sehingga kita harus menyesuaikannya dengan tujuan yang ada
3. Visualisasi data dapat memberikan informasi lebih lengkap dibandingkan dengan data aslinya

4.2 Saran

Meskipun memiliki berbagai bentuk, visualisasi data yang telah dilakukan masih batasan variabel dan batasan teori. Namun, variasi bentuk yang ditampilkan bisa lebih banyak bergantung pada keinginan pembuat. Data yang digunakan juga masih harus diolah lebih lanjut. Hal tersebut memperpanjang proses visualisasi data. Apalagi jika kita mengambil data secara *real-time*. Kita bisa membentuk tim atau kelompok, sehingga proses visualisasi data beserta pengolahannya akan lebih singkat.

Visualisasi data yang telah tersedia biasanya juga terbatas pada satu variasi warna ataupun bentuk, sehingga pembaca akan kurang nyaman jika memiliki gangguan penglihatan dan sejenisnya. Meskipun sudah tersedia berbagai macam bentuk beserta animasinya, kita masih perlu mengeksplorasi proses visualisasi data sehingga dapat menampilkan yang dinamis sesuai keinginan pembaca.

DAFTAR PUSTAKA

- IT Salary Survey December 2020. Retrieved April 11, 2021, from
<https://www.asdcode.de/2021/01/it-salary-survey-december-2020.html>
- R. Retrieved April 11, 2021, from [https://id.wikipedia.org/wiki/R_\(bahasa_pemrograman\)](https://id.wikipedia.org/wiki/R_(bahasa_pemrograman))
- RStudio. (n.d.). Retrieved April 11, 2021, from <https://id.wikipedia.org/wiki/RStudio>
- Visualisasi data. Retrieved April 11, 2021, from
https://id.wikipedia.org/wiki/Visualisasi_data